

Transition states in protein folding kinetics: Modeling Φ -values of small β -sheet proteins

Thomas R. Weikl

Max Planck Institute of Colloids and Interfaces, Department of
Theory and Bio-Systems, 14424 Potsdam, Germany

Abstract

Small single-domain proteins often exhibit only a single free-energy barrier, or transition state, between the denatured and the native state. The folding kinetics of these proteins is usually explored via mutational analysis. A central question is which structural information on the transition state can be derived from the mutational data. In this article, we model and structurally interpret mutational Φ -values for two small β -sheet proteins, the PIN and the FBP WW domain. The native structure of these WW domains comprises two β -hairpins that form a three-stranded β -sheet. In our model, we assume that the transition state consists of two conformations in which either one of the hairpins is formed. Such a transition state has been recently observed in Molecular Dynamics folding-unfolding simulations of a small designed three-stranded β -sheet protein. We obtain good agreement with the experimental data (i) by splitting up the mutation-induced free-energy changes into terms for the two hairpins and for the small hydrophobic core of the proteins, and (ii) by fitting a single parameter, the relative degree to which hairpin 1 and 2 are formed in the transition state. The model helps to understand how mutations affect the folding kinetics of WW domains, and captures also negative Φ -values that have been difficult to interpret.

Introduction

How proteins fold into their native 3-dimensional structure remains an intriguing question. Given the vast number of unfolded protein conformations, Cyrus

Levinthal argued in 1968 [1, 2] that proteins are guided to their native structure by a sequence of folding intermediates. In the following decades, experimentalists focused on detecting and characterizing metastable intermediates with a variety of methods [3]. While such folding intermediates continue to be of considerable interest [4, 5], the view that proteins have to fold in sequential pathways from intermediate to intermediate, now known as ‘old view’ [6, 7], changed in the ’90s when statistical-mechanical models demonstrated that fast and efficient folding can also be achieved on funnel energy landscapes that are smoothly biased towards the native state and do not exhibit metastable intermediates [8, 9]. The paradigmatic proteins of this ‘new view’ are two-state proteins, first discovered in 1991 [10]. Two-state proteins fold from the denatured state to the native state without experimentally detectable intermediate states. Since then, many small-single domain proteins have been shown to fold in two-state kinetics [11–13].

The folding dynamics of two-state proteins is thought to be dominated by a single free-energy barrier, or transition state, between the denatured and native state. This transition state of the protein folding reaction is an unstable, short-lived state and cannot be observed directly. Instead, the dynamics of two-state proteins is often explored via mutational analysis [14–33]. In such an analysis, a large number of mostly single-residue mutants of a protein is generated. For each mutant, the effect of the mutation on the folding dynamics is usually quantified by its Φ -value [12, 34]

$$\Phi = \frac{RT \ln(k_{\text{wt}}/k_{\text{mut}})}{\Delta G_N} \quad (1)$$

Here, k_{wt} is the folding rate for the wildtype protein, k_{mut} is the folding rate for the mutant protein, and ΔG_N is the change of the protein stability induced by the mutation. The stability G_N of a protein is the free energy difference between the denatured state D and the native state N . In classical transition-state theory, the folding rate of a two-state protein is proportional to $\exp[-G_T/RT]$, where G_T is the free energy difference from the denatured state to the transition state. It is usually assumed that the prefactor of this proportionality relation does not depend on the mutation. In this notation, Φ -values have the form

$$\Phi = \frac{\Delta G_T}{\Delta G_N} \quad (2)$$

where ΔG_T is the mutation-induced change of the free-energy barrier G_T .

The mutational Φ -value data for a protein provide indirect information on its folding dynamics and, therefore, have attracted considerable theoretical interest. The central question is: Which transition transition-state structures and free-energy perturbations are consistent with the experimentally measured Φ -values?

In this article, we model Φ -values from detailed mutational analyses [25, 29, 33] of two small β -sheet proteins, the FBP and PIN WW domains. The native structure of these proteins consists of two hairpins forming a three-stranded sheet [35, 36] (see Fig. 1). The design principles [37, 38] and folding kinetics [25, 29, 33, 39–47] of WW domains and other three-stranded β -sheet proteins have been studied extensively. Because of their small size and abundance as protein domains, WW domains are important model systems for understanding β -sheet folding and stability.

Molecular Dynamics (MD) simulations with atomistic models are computationally demanding and in general do not allow direct calculations of folding rates and Φ -values. With additional assumptions on transition states or Φ -values, transition-state conformations have been extracted from MD unfolding trajectories at elevated temperatures [48–50] or constructed from MD simulations that use Φ -values as restraints [51, 52].¹ However, for a small, designed three-stranded β -sheet protein, beta3s, transition-state conformations [60] and Φ -values [61] have been more rigorously determined from extensive equilibrium folding-unfolding MD simulations. The native structure of beta3s is similar to the structure of WW domains, with two β -hairpins forming an antiparallel three-stranded β -sheet. Rao et al. [60] performed four MD simulations of beta3s at the temperature 330 K with a total length 12.6 μ s, and observed 72 folding and 73 unfolding events. By identifying clusters of structurally similar conformations that have the probability $p_{\text{fold}} = 0.5$ to fold [62–64], and the same probability to unfold, Rao et al. obtained a transition-state ensemble for beta3s that is “characterized by the presence of one of the two native hairpins formed while the rest of the peptide is mainly unstructured” [60]. The two β -hairpins of beta3s thus appear to be cooperative substructures that are either fully structured or unstructured in the transition state.

Here, we show that a statistical-mechanical model with a beta3s-like transition-state ensemble in which either hairpin 1 or hairpin 2 are formed leads to an overall consistent interpretation of experimental Φ -values for the FBP and PIN WW domains. In this model, mutations can either affect hairpin 1, hairpin 2, or the small hydrophobic core of the WW domains, which is not yet structured in the transition state. The general form of Φ -values in this model is

$$\Phi = \frac{\Delta G_T}{\Delta G_N} = \frac{\chi_1 \Delta G_1 + \chi_2 \Delta G_2}{\Delta G_N} \quad (3)$$

where χ_1 is the probability, or fraction, of the transition-state conformation in which hairpin 1 is formed, and $\chi_2 = 1 - \chi_1$ is the probability of the transition-

¹In statistical-mechanical or Go-type models with simplified energy landscapes, in contrast, folding rates and stabilities for wildtype and mutants can be easily calculated [53–59]. However, the lack of atomistic detail in these models appears to make it difficult to reproduce detailed mutational data.

state conformation with hairpin 2 formed. The mutation-induced changes of the free energy difference between the two transition-state conformations and the denatured state are denoted by ΔG_1 and ΔG_2 . The model has just two structural parameters, χ_1 and χ_2 , which are obtained from a comparison with the experimental data. Different Φ -values for different mutations simply arise from different ‘free-energy signatures’ ΔG_1 , ΔG_2 , and ΔG_N of the mutations.

In particular, the model reproduces the negative Φ -value for the mutation L36A of the FBP WW domain. The mutation destabilizes the native state ($\Delta G_N > 0$), but stabilizes hairpin 2 ($\Delta G_2 < 0$), according to calculations with the empirical force field FOLD-X [65,66]. This leads to a negative Φ -value in eq. (3) since ΔG_1 equals 0 for this mutation. In general, ‘nonclassical’ Φ -values, i.e. Φ -values that are negative or larger than 1, are obtained in the model if mutations stabilize some structural elements, but destabilize others. The mutation L36A of the FBP WW domain, for example, stabilizes hairpin 2, but destabilizes the hydrophobic core.

Nonclassical Φ -values have been difficult to interpret in the traditional interpretation. In this interpretation, a Φ -value is taken to indicate the degree of structure formation of the mutated residue in the transition-state ensemble T [12]. A Φ -value of 1 is interpreted to indicate that the residue has a native-like structure in T, since the mutation shifts the free energy of the transition state T by the same amount as the free energy of the native state N. A Φ -value of 0 is interpreted to indicate that the residue is as unstructured in T as in the denatured state D, since the mutation does not shift the free-energy difference between these two states. Φ -values between 0 and 1 are typically taken to indicate partial native-like structure in T. For a protein with M residues, the traditional interpretation thus implies M structural parameters, the degrees of structure formation of all residues. In contrast, the model presented here has just a single independent parameter, the relative degree to which hairpin 1 and 2 are populated in T. Since degrees of structure formation have to be between 0 (‘denatured-like’) and 1 (‘native-like’), the traditional interpretation can not explain nonclassical Φ -values smaller than 0 or larger than 1. In the present model, nonclassical Φ -values arise from substructural free-changes contributions of different sign (see above).

We have recently suggested a related, novel model for Φ -values of mutations in α -helices of a protein [67,68]. The model is based on cooperative helix formation and on splitting mutation-induced free energy changes in helices into secondary and tertiary terms [68]. The two structural model parameters are the degrees of secondary and tertiary structure formation of the helix in the transition state. For several well-characterized helices [68], fitting these two parameters to mutational data leads to a consistent, structural interpretation of the Φ -values. The general

conclusion from our helix model and the model for small β -sheet proteins presented here is that a consistent structural interpretation of Φ -values (i) requires to split up mutation-induced stability changes into free-energy contributions from different substructural elements of a protein, and (ii) can be obtained with few parameters that characterize the degree of structure formation of cooperative elements such as α -helices and β -hairpins in the transition-state ensemble.

Model

The central assumption of our model is that each of the hairpins is either fully formed or not formed in the transition-state ensemble of the protein. The model has then four states: the denatured state D in which none of the hairpins is formed, a transition-state conformation in which only hairpin 1 is formed, a transition-state conformation in which only hairpin 2 is formed, and the native state with both hairpins formed. The energy landscape can be characterized by three free-energy differences: The free-energy difference G_N of the native state and the free-energy differences G_1 and G_2 of the transition-state conformations with respect to the denatured state (see Fig. 2).

The folding kinetics is described by the master equation

$$\frac{dP_n(t)}{dt} = \sum_{m \neq n} [w_{nm}P_m(t) - w_{mn}P_n(t)], \quad (4)$$

which gives the time evolution of the probability $P_n(t)$ that the protein is in state n at time t . Here, w_{nm} is the transition rate from state m to n , defined by

$$w_{nm} = \frac{1}{t_o} (1 + e^{G_n - G_m})^{-1} \quad (5)$$

provided the states n and m are connected via a single step in which only a single hairpin folds or unfolds [69]. For other transitions, i.e. for the direct transition from the denatured state to the native state, and vice versa, the transition rates are zero. Here, t_o is a reference time scale. The transition rates defined above obey detailed balance $w_{nm}P_m^e = w_{mn}P_n^e$ where $P_n^e \sim \exp[-G_n/(RT)]$ is the equilibrium weight for the state n . Detailed balance ensures that the system ultimately reaches thermal equilibrium.

The master equation of this four-state model can be solved exactly (see Appendix). For high transition-state barriers $G_1 \gg RT$ and $G_2 \gg RT$ and a stable native

state with $G_N \ll -RT$, the folding rate is given by

$$k(G_1, G_2) \simeq \frac{1}{2} (e^{-G_1/RT} + e^{-G_2/RT}) \quad (6)$$

in units of $1/t_o$. The folding rate k simply is the sum of the rates for the two possible folding routes on which either hairpin 1 or hairpin 2 forms first. The factor $\frac{1}{2}$ in the equation above arises because a molecule, after reaching one of the barrier states 1 or 2, either proceeds to N or returns to D , with almost equal probability.

Mutations correspond to the perturbations of the free energy landscape. A mutation therefore can be characterized by the free energy changes ΔG_1 , ΔG_2 , and ΔG_N . The folding rate of the mutant then is $k_{\text{mut}} \equiv k(G_1 + \Delta G_1, G_2 + \Delta G_2)$. For small perturbations ΔG_1 and ΔG_2 , a Taylor expansion of $\ln k_{\text{wt}} \equiv \ln k$ to first order leads to

$$\ln k_{\text{mut}} - \ln k_{\text{wt}} \simeq \frac{\partial \ln k}{\partial G_1} \Delta G_1 + \frac{\partial \ln k}{\partial G_2} \Delta G_2 = -\frac{1}{RT} (\chi_1 \Delta G_1 + \chi_2 \Delta G_2) \quad (7)$$

with

$$\chi_1 \equiv \frac{e^{-G_1/RT}}{e^{-G_1/RT} + e^{-G_2/RT}} \quad \text{and} \quad \chi_2 \equiv \frac{e^{-G_2/RT}}{e^{-G_1/RT} + e^{-G_2/RT}} \quad (8)$$

The two parameters χ_1 and χ_2 quantify the extent to which the transition-state conformation 1 and the transition-state conformation 2 are populated in the transition-state ensemble. From the Φ -value definition (1) and eq. (7), we obtain the general form of Φ -values given in eq. (3).

Results

FBP WW domain

We first consider the FBP WW domain. Petrovich et al. [33] have performed an extensive mutational analysis of the folding kinetics. The Φ -values and stability changes ΔG_N for the considered mutations are summarized in Table 1, together with an assessment which structural elements are affected by the mutations. This assessment is based on the contact matrix of the FBP WW domain shown in Fig. 3. A black dot at position (i, j) of this matrix indicates that the two amino acids i and j are in contact, i.e. that the distance between any of their non-hydrogen atoms is smaller than the cutoff distance 4 Å. Since the contact matrix is symmetric, only one half is represented in Fig. 3. The two contact clusters in the matrix correspond

to hairpin 1 and hairpin 2 of the FBP WW domain. The remaining contacts largely correspond to contacts of hydrophobic amino acids, the small ‘hydrophobic core’ of the protein. About half of the mutations performed by Petrovich et al. affect only either hairpin 1 or hairpin 2. The mutation E7A of amino acid 7, for example, affects the contacts (7, 22), (7, 23), and (7, 24), which are all located in hairpin 1 (see contact map in Fig. 3). The remaining mutations also affect the hydrophobic core, or both hairpins. The mutation Y21A, for example affects the contacts (8, 21) and (9, 21) in hairpin 1, and the contacts (21, 26), (21, 27), and (21, 28) in hairpin 2.

To test our model, we first consider all mutations that affect only one of the hairpins. The model predicts that all mutations that affect only hairpin 1 should have the same Φ -value χ_1 , and all mutations that affect only hairpin 2 the same Φ -value χ_2 . This is a direct consequence of eq. (3). For mutations that affect only hairpin 1, for example, we have $\Delta G_2 = 0$ since the mutations don’t shift the stability of hairpin 2, and $\Delta G_N = \Delta G_1$ since they also don’t affect the hydrophobic core. Eq. (3) then results in $\Phi = \chi_1$ for these mutations. The Φ -values for the ten mutations that only affect hairpin 1 are plotted in Fig. 4. Except for one clear outlier,² all Φ -values are centered around the value 0.8, mostly within experimental errors. The mean value of these nine Φ -values (dashed line in Fig. 4) leads to the estimated $\chi_1 = 0.81 \pm 0.06$. The error here is estimated as error of the sample mean. The standard deviation of the Φ -values from the mean value is 0.18. The four Φ -values for mutations that affect only hairpin 2 range from 0.08 to 0.39 (see Table 1), with mean value $\chi_2 = 0.30 \pm 0.08$ and standard deviation 0.16. For both sets of mutations, we thus obtain good agreement with the model. In addition, the sum of the above estimated values for the model parameters χ_1 and χ_2 is close to 1, within the error bounds, which is an additional consistency requirement of the model. The two parameters χ_1 and χ_2 are the fractions to which the two transition-state conformations with either hairpin 1 or hairpin 2 formed are populated. These fractions sum up to 1 since the protein has to take one of the possible routes in the model.

To include other mutations in the model, we have to estimate the impact of these mutations on the stability of the different structural elements they affect (hairpin 1, hairpin 2, or the hydrophobic core). We use FOLD-X here, a molecular modeling program for the prediction of mutation-induced stability changes [65,66]. The

²The data point for the mutation T9A can be confirmed as outlier, e.g., with the Grubb’s test [70] at the standard significance level of 5 %. For a set of 10 data points as here, a value of x is an outlier for $z \equiv (x - \bar{x})/\text{SD} > 2.29$ where \bar{x} is the sample mean, and SD the standard deviation. For the mutation T9A with Φ -value -0.09 , the z -value 2.43 exceeds the critical value 2.29.

FOLD-X force field includes terms for backbone and sidechain entropies, which have been weighted against other terms using experimental data from mutational stability analyses. FOLD-X has been tested on a set of 1088 point mutants and reproduces the stability changes of 1030 of these mutants with a correlation coefficient of 0.83 and a standard deviation of 0.81 kcal/mol [65]. With FOLD-X, we calculate the mutation-induced stability changes ΔG_N for the whole FBP WW domain, and the stability changes ΔG_1 and ΔG_2 of hairpin 1 and 2, depending on whether the mutation affects these hairpins. To calculate ΔG_1 and ΔG_2 , we simply ‘cut out’ these hairpins from the PDB structure and estimate the stability of the wildtype and mutant hairpins with FOLD-X (see caption of Table 2 for details). The resulting data are summarized in Table 2. The calculated stability changes ΔG_N can be directly compared to the experimentally measured stability changes $\Delta G_{N,\text{exp}}$. We include here only mutations in the model for which the FOLD-X predicted stability changes ΔG_N do not differ by more than a factor 2 from the experimental stability changes $\Delta G_{N,\text{exp}}$. For other mutations, the force-field calculations are unreliable. In Table 2, the calculated stability changes for these mutations are shown in brackets.

The mutations in Table 2 affect two of the structural elements: The mutations W8F and T13A affect hairpin 1 and the hydrophobic core. For these mutations, we have $\Delta G_2 = 0$, and $\Phi = \chi_1 \Delta G_1 / \Delta G_N$ according to Eq. (3). The mutation Y21A affects both hairpins, hence $\Phi = (\chi_1 \Delta G_1 + \chi_2 \Delta G_2) / (\Delta G_1 + \Delta G_2)$. Finally, the mutations T29G, W30A, and L36V affect hairpin 2 and the hydrophobic core. Therefore, we have $\Delta G_1 = 0$ for these mutations, and $\Phi = \chi_2 \Delta G_2 / \Delta G_N$.

Let us now consider the set of 20 mutations that consists of these 6 mutations that affect two structural elements and the 14 mutations that affect either only hairpin 1 or only hairpin 2. Our model has two parameters, χ_1 and χ_2 . However, since $\chi_1 + \chi_2 = 1$, there is only one independent parameter. We determine this parameter from a least-square fit between the theoretical Φ -value formula given in eq. 3 and the experimental Φ -values and obtain the values $\chi_1 = 0.77 \pm 0.05$ and $\chi_2 = 0.23 \pm 0.05$, see Fig. 5.

PIN WW domain

Mutational analyses of the PIN WW domain’s folding kinetics have been performed by Jäger et al. [25] and Deechongkit et al. [29]. While Jäger et al. have considered standard single-site amino-acid replacements, Deechongkit et al. synthesized amid-to-ester mutants that specifically perturb backbone H-bonds. The experimental Φ -values and stability changes $\Delta G_{N,\text{exp}}$ for these mutations are summarized in Table

3. The synthetic amino acids in the mutations of Deechongkit et al. are denoted by lowercase greek letters (last six lines in Table 3). Since these mutations perturb the backbone H-bonds, they only affect either hairpin 1 or hairpin 2, which is indicated in the last column in Table 3. For the mutations considered by Jäger et al., the affected structural elements are again assessed based on the contact map shown in Fig. 3. We consider here only mutations with stability changes $\Delta G_{N,\text{exp}} > 0.8$ kcal/mol. Φ -values of mutations that cause significantly smaller stability changes are often considered as unreliable [30, 71, 72] (see also Discussion).

Seven mutations in Table 3 affect only hairpin 1 of the PIN WW domain. The mean value of the Φ -values for these mutations leads to the estimate $\chi_1 = 0.69 \pm 0.05$. The standard deviation of the Φ -values from the mean is 0.12, which is comparable to the experimental errors. The four Φ -values of the mutations that affect only hairpin 2 have the mean value $\chi_2 = 0.36 \pm 0.05$ and the standard deviation 0.10. In agreement with our model, these estimates for χ_1 and χ_2 again add up to 1, within the statistical errors. In an alternative approach, the values of χ_1 and χ_2 can be obtained from a least-square fit between theoretical and experimental Φ -values (see Fig. 6). From the fit, we obtain $\chi_1 = 0.67 \pm 0.05$ and $\chi_2 = 1 - \chi_1 = 0.33 \pm 0.05$, and a Pearson correlation coefficient of 0.85 between theoretical and experimental Φ -values.

We do not include mutations that affect more than one structural element here since the stability changes estimated with FOLD-X appear to be unreliable. For four of the five mutants, the calculated stability changes ΔG_N differ by significantly more a factor 2 from experimental values $\Delta G_{N,\text{exp}}$ (data not shown). The stabilities for the PIN WW domain mutants may be more difficult to calculate since they involve a larger range of amino acids, compared to the FBP WW mutants that mostly involve changes to the small amino acids Alanine or Glycine, which can be modeled via simple truncation of sidechains prior to the FOLD-X calculations.

Discussion and Conclusions

We have modeled Φ -values from extensive mutational analyses of two WW domains based on the central assumption that the transition state ensemble of these proteins consists of two substates in which either hairpin 1 or hairpin 2 are formed. The structural information obtained from the mutational data by fitting a single model parameter is that the transition state ensemble of the FBP WW domains consists to roughly $\frac{3}{4}$ of substate 1 with hairpin 1 formed, and to $\frac{1}{4}$ of substate 2 with hairpin 2 formed. The transition state ensemble of the PIN WW domain consists

to roughly $\frac{2}{3}$ of substate 1, and to $\frac{1}{3}$ of substate 2, according to the model.

In the model, the magnitude of a Φ -value depends on which structural elements are affected, and on the mutation-induced free energy changes of these elements. The mutation E7A of the FBP WW domain, for example, has a relatively large Φ -value since this mutation only affects hairpin 1, which is structured in the dominant substate 1 of the transition state ensemble, whereas the mutation W8F has a relatively small Φ -value since the mutation mainly affects the free energy of the small hydrophobic core, which is not yet formed in the transition state. The model also reproduces the negative Φ -value of the mutation L36A, which results from different signs of the mutations-induced free energy changes ΔG_1 and ΔG_N in Table 2. According to the free-energy calculations with FOLD-X, the mutation stabilizes hairpin 1 ($\Delta G_1 < 0$), but has an overall destabilizing effect ($\Delta G_N > 0$) since it destabilizes the hydrophobic core.

The deviations between experimental and theoretical Φ -values are within reasonable errors. It has been recently suggested that experimental errors for Φ -values may be underestimated since it is usually assumed that the errors in the measured free energy changes of the transition state and the folded state are independent, which is not the case [73]. In case of the PIN WW domain, we have only considered mutations with stability changes $\Delta G_N > 0.8$ kcal/mol. For mutations that induce significantly smaller stability changes, experimental errors in ΔG_N may lead to large errors in Φ -values since ΔG_N constitutes the denominator of the Φ -value defined in eq. (1).

However, the large Φ -values up to 1.8 for three mutations with small stability changes in the loop of hairpin 1 of the PIN WW domain [25], which have not been considered here, may also result from structural rearrangements. Jäger et al. [25] have suggested a five-state model with two consecutive transition states. In the first transition state, only the loop of hairpin 1 is formed. Nonclassical Φ -values greater than 1 are obtained in this model for mutations that are assumed to shift the free energy of the loop by a larger amount than the free energy of the native state. With the same assumption, large nonclassical Φ -values in the loop of hairpin 1 are also obtained in the four-state model presented here. For $\chi_1 = 0.67$, for example, a Φ -value of 1.8 is obtained for a mutation in this loop with $\Delta G_1 = 2.7\Delta G_N$, according to eq. (3), since hairpin 2 and, thus, ΔG_2 are not affected by this mutation. Such a situation may result from a structural rearrangement between the transition-state conformation with hairpin 1 formed and the native state. The structural rearrangement may affect the sidechains in the loop, but should not affect the backbone hydrogen bonds since the Φ -values for the amide-to-ester mutations S16 σ , R17 ρ , and S19 σ in this loop are between

0.70 and 0.83 (see Table 3). Within the experimental and statistical errors, these Φ -values are close to $\chi_1 = 0.67$, which is the expected Φ -value for mutations with $\Delta G_1 = \Delta G_N$.

Appendix: Exact solution of the master equation

The master equation (4) can be written in the matrix form:

$$\frac{d\mathbf{P}(t)}{dt} = -\mathbf{W}\mathbf{P}(t) \quad (9)$$

The elements of the vector $\mathbf{P}(t)$ are the probabilities $P_n(t)$ that the protein is in state n at time t , and the matrix elements of \mathbf{W} are given by

$$W_{nm} = -w_{nm} \quad \text{for } n \neq m; \quad W_{nn} = \sum_{m \neq n} w_{mn}. \quad (10)$$

For the model with four states considered here, the matrix \mathbf{W} is given by

$$\mathbf{W} = \frac{1}{t_o} \begin{pmatrix} \frac{1}{1+e^{g_1}} + \frac{1}{1+e^{g_2}} & -\frac{1}{1+e^{-g_1}} & -\frac{1}{1+e^{-g_2}} & 0 \\ -\frac{1}{1+e^{g_1}} & \frac{1}{1+e^{-g_1}} + \frac{1}{1+e^{g_N-g_1}} & 0 & -\frac{1}{1+e^{g_1-g_N}} \\ -\frac{1}{1+e^{g_2}} & 0 & \frac{1}{1+e^{-g_2}} + \frac{1}{1+e^{g_N-g_2}} & -\frac{1}{1+e^{g_2-g_N}} \\ 0 & -\frac{1}{1+e^{g_N-g_1}} & -\frac{1}{1+e^{g_N-g_2}} & \frac{1}{1+e^{g_1-g_N}} + \frac{1}{1+e^{g_2-g_N}} \end{pmatrix}$$

To simplify the notation, we have used here dimensionless free-energy differences $g_i \equiv G_i/RT$ ($i = 1, 2$, or N) of the partially folded states 1 and 2 and the native state N with respect to the denatured state.

The general solution $\mathbf{P}(t)$ of the master equation can be expressed in terms of the eigenvalues λ and eigenvectors \mathbf{Y}_λ of the matrix \mathbf{W} :

$$\mathbf{P}(t) = \sum_{\lambda} c_{\lambda} \mathbf{Y}_{\lambda} \exp[-\lambda t] \quad (11)$$

The prefactors c_{λ} in this general solution depend on the initial conditions at time $t = 0$. For the 4×4 matrix above, the 4 eigenvalues are given by $\lambda = 0, 1 - q, 1 + q$, and 2, in units of $1/t_o$, with

$$q \equiv \frac{1 - e^{g_N-g_1-g_2}}{\sqrt{(1+e^{-g_1})(1+e^{-g_2})(1+e^{g_N-g_1})(1+e^{g_N-g_2})}} \quad (12)$$

Since we have $-1 < q < 1$, the three nonzero eigenvalues are positive and describe the relaxation to the equilibrium state of the model (see eq. (11)). The equilibrium state simply is $c_o \mathbf{Y}_o$ where \mathbf{Y}_o is the eigenvector with eigenvalue 0.

This model exhibits two-state folding kinetics under two conditions. First, the native state has to be stable, i.e. the free energy g_N of the native state must be significantly smaller than the free energies of the other three states. Second, the free energy differences g_1 and g_2 between the intermediate states and the denatured have to be significantly larger than RT . The partially folded states then constitute the transition-state ensemble. Under these two conditions, the three Boltzmann weights $e^{g_N - g_1 - g_2}$, $e^{g_N - g_1}$, and $e^{g_N - g_2}$ in eq. (12) are much smaller than 1, and also much smaller than e^{-g_1} and e^{-g_2} , which leads to

$$q \simeq \frac{1}{\sqrt{(1 + e^{-g_1})(1 + e^{-g_2})}} \quad (13)$$

For large barrier energies g_1 and g_2 , we have $e^{-g_1} \ll 1$ and $e^{-g_2} \ll 1$, and therefore $(1 + e^{-g_1})(1 + e^{-g_2}) \simeq (1 + e^{-g_1} + e^{-g_2})$. If we now use the expansion $(1 + x)^{-1/2} \simeq 1 - x/2$ with $x = e^{-g_1} + e^{-g_2} \ll 1$, the smallest nonzero relaxation rate, or folding rate, $k \equiv 1 - q$ is given by eq. (6), i.e. by $k \simeq \frac{1}{2}(e^{-g_1} + e^{-g_2})$ in the notation used in this appendix. The folding rate k is much smaller than the other two relaxation rates $1 + q$ and 2, which corresponds to an initial ‘burst phase’.

References

- [1] Levinthal, C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* 65:44–45.
- [2] Levinthal, C. 1969. How to fold gracefully. *In* Mössbauer spectroscopy in biological systems. University of Illinois Bulletin. 67:22–24.
- [3] Baldwin, R. L. 1999. Protein folding from 1961 to 1982. *Nat. Struct. Biol.* 6:814–817.
- [4] Baldwin, R. L., and G. D. Rose. 1999. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends biochem. sci.* 24:77–83.
- [5] Cecconi, C., E. A. Shank, C. Bustamante, and S. Marqusee. 2005. Direct observation of the three-state folding of a single protein molecule. *Science*. 309:2057–2060.
- [6] Baldwin, R. L. 1994. Matching speed and stability. *Nature*. 369:183–184.
- [7] Matthews, C. R. 1993. Pathways of protein folding. *Annu. Rev. Biochem.* 62:653–683.

- [8] Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.
- [9] Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
- [10] Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor-2. 1. Evidence for a two-state transition. *Biochemistry.* 30:10428–10435.
- [11] Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:R81–R91.
- [12] Fersht, A. R. 1999. Structure and mechanism in protein science. W. H. Freeman, New York.
- [13] Grantcharova, V., E. J. Alm, D. Baker, and A. L. Horwich. 2001. Mechanisms of protein folding. *Curr. Opin. Struct. Biol.* 11:70–82.
- [14] Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.
- [15] Villegas, V., J. C. Martinez, F. X. Aviles, and L. Serrano. 1998. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* 283:1027–1036.
- [16] Chiti, F., N. Taddei, P. M. White, M. Bucciantini, F. Magherini, M. Stefani, and C. M. Dobson. 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* 6:1005–1009.
- [17] Ternström, T., U. Mayor, M. Akke, and M. Oliveberg. 1999. From snapshot to movie: Φ analysis of protein folding transition states taken one step further. *Proc. Natl. Acad. Sci. USA.* 96:14854–14859.
- [18] Kragelund, B. B., P. Osmark, T. B. Neergaard, J. Schiodt, K. Kristiansen, J. Knudsen, and F. M. Poulsen. 1999. The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nat. Struct. Biol.* 9:594–601.
- [19] Martinez, J. C., and L. Serrano. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6:1010–1016.

- [20] Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski, and D. Baker. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6:1016–1024.
- [21] Fulton, K. F., E. R. G. Main, V. Daggett, and S. E. Jackson. 1999. Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* 291:445–461.
- [22] Hamill, S. J., A. Steward, and J. Clarke. 2000. The folding of an immunoglobulin-like greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* 297:165–178.
- [23] Kim, D. E., C. Fisher, and D. Baker. 2000. A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* 298:971–984.
- [24] McCallister, E. L., E. Alm, and D. Baker. 2000. Critical role of β -hairpin formation in protein G folding. *Nat. Struct. Biol.* 7:669–673.
- [25] Jäger, M., H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele. 2001. The folding mechanism of a β -sheet: The WW domain. *J. Mol. Biol.* 311:373–393.
- [26] Otzen, D. E., and M. Oliveberg. 2002. Conformational plasticity in folding of the split β - α - β protein S6: Evidence for burst-phase disruption of the native state. *J. Mol. Biol.* 317:613–627.
- [27] Northey, J. G. B., A. A. Di Nardo, and A. R. Davidson. 2002. Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* 9:126–130.
- [28] Gianni, S., N. R. Guydosh, F. Khan, T. D. Caldas, U. Mayor, G. W. N. White, M. L. DeMarco, V. Daggett, and A. R. Fersht. 2004. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA.* 100:13286–13291.
- [29] Deechongkit, S., H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, and J. W. Kelly. 2004. Context-dependent contributions of backbone hydrogen bonding to β -sheet folding energetics. *Nature.* 430:101–105.
- [30] Garcia-Mira, M. M., D. Böhringer, and F. X. Schmid. 2004. The folding transition state of the cold shock protein is strongly polarized. *J. Mol. Biol.* 339:555–569.
- [31] Anil, B., S. Sato, J. H. Cho, and D. P. Raleigh. 2005. Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J. Mol. Biol.* 354:693–705.

- [32] Wilson, C. J., and P. Wittung-Stafshede. 2005. Snapshots of a dynamic folding nucleus in zinc-substituted *Pseudomonas aeruginosa* azurin. *Biochemistry*. 44:10054–10062.
- [33] Petrovich, M., A. L. Jonsson, N. Ferguson, V. Daggett, and A. R. Fersht. 2006. Φ -analysis at the experimental limits: Mechanism of β -hairpin formation. *J. Mol. Biol.* 360:865–881.
- [34] Matouschek, A., J. T. Kellis, L. Serrano, and A. R. Fersht. 1989. Mapping the transition state and pathway of protein folding by protein engineering. *Nature*. 340:122–126.
- [35] Macias, M. J., V. Gervais, C. Civera, and H. Oschkinat. 2000. Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* 7:375–379.
- [36] Ranganathan, R., K. P. Lu, T. Hunter, and J. P. Noel. 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell*. 89:875–886.
- [37] Socolich, M., S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner, and R. Ranganathan. 2005. Evolutionary information for specifying a protein fold. *Nature*. 437:512–518.
- [38] Fernandez-Escamilla, A. M., S. Ventura, L. Serrano, and M. A. Jimenez. 2006. Design and NMR conformational study of a β -sheet peptide based on betanova and WW domains. *Protein Sci.* 15:2278–2289.
- [39] Ferguson, N., J. R. Pires, F. Toepert, C. M. Johnson, Y. P. Pan, R. Volkmer-Engert, J. Schneider-Mergener, V. Daggett, H. Oschkinat, and A. Fersht. 2002. Using flexible loop mimetics to extend Φ -value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. USA*. 98:13008–13013.
- [40] Nguyen, H., M. Jäger, A. Moretto, M. Gruebele, and J. W. Kelly. 2003. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA*. 100:3948–3953.
- [41] Deechongkit, S., H. Nguyen, M. Jager, E. T. Powers, M. Gruebele, and J. W. Kelly. 2006. β -sheet folding mechanisms from perturbation energetics. *Curr. Opin. Struct. Biol.* 16:94–101.
- [42] Jäger, M., Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, and J. W. Kelly. 2006. Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci. USA*. 103:10648–10653.

- [43] Bursulaya, B. D., and C. L. Brooks III. 1999. Folding free energy surface of a three-stranded β -sheet protein. *J. Am. Chem. Soc.* 121:9947–9951.
- [44] Ferrara, P., and A. Caffisch. 2000. Folding simulations of a three-stranded antiparallel β -sheet peptide. *Proc. Natl. Acad. Sci. USA.* 97:10780–10785.
- [45] Davis, R., C. M. Dobson, and M. Vendruscolo. 2002. Determination of the structures of distinct transition state ensembles for a β -sheet peptide with parallel folding pathways. *J. Chem. Phys.* 117:9510–9517.
- [46] Karanickolas, J., and C. L. Brooks III. 2004. Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain. *Proc. Natl. Acad. Sci. USA.* 101:3432–3437.
- [47] Bruscolini, P., and F. Cecconi. 2005. Analysis of PIN1 WW domain through a simple statistical mechanics model. *Biophys. Chem.* 115:153–158.
- [48] Li, A., and V. Daggett. 1996. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by Molecular Dynamics simulations. *J. Mol. Biol.* 257:412–429.
- [49] Day, R., and V. Daggett. 2005. Sensitivity of the folding/unfolding transition state ensemble of chymotrypsin inhibitor 2 to changes in temperature and solvent. *Protein Sci.* 14:1242–1252.
- [50] Duan, J., and L. Nilsson. 2005. Thermal unfolding simulations of a multimeric protein – Transition state and unfolding pathways. *Proteins.* 59:170–182.
- [51] Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2001. Three key residues form a critical contact network in a protein folding transition state. *Nature.* 409:641–645.
- [52] Paci, E., M. Vendruscolo, C. M. Dobson, and M. Karplus. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* 324:151–163.
- [53] Alm, E., and D. Baker. 1999. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA.* 96:11305–11310.
- [54] Muñoz, V., and W. A. Eaton. 1999. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA.* 96:11311–11316.

- [55] Galzitskaya, O. V., and A. V. Finkelstein. 1999. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA*. 96:11299–11304.
- [56] Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
- [57] Guerois, R., and L. Serrano. 2000. The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* 304:967–982.
- [58] Alm, E., A. V. Morozov, T. Kortemme, and D. Baker. 2002. Simple physical models connect theory and experiment in protein folding kinetics. *J. Mol. Biol.* 322:463–476.
- [59] Kameda, T. 2000. Importance of sequence specificity for predicting protein folding pathways: Perturbed Gaussian chain model. *Proteins*. 53:616–628.
- [60] Rao, F., G. Settanni, E. Guarnera, and A. Caffisch. 2005. Estimation of protein folding probability from equilibrium simulations. *J. Chem. Phys.* 122:184901.
- [61] Settanni, G., F. Rao, and A. Caffisch. 2005. Φ -value analysis by molecular dynamics simulations of reversible folding. *Proc. Natl. Acad. Sci. USA*. 102:628–633.
- [62] Du, R., V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich. 1998. On the transition coordinate for protein folding. *J. Chem. Phys.* 108:334–350.
- [63] Hummer, G. 2004. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* 120:516–523.
- [64] Snow, C. D., Y. M. Rhee, and V. S. Pande. 2006. Kinetic definition of protein folding transition state ensembles and reaction coordinates. *Biophys J.* 91:14–24.
- [65] Guerois, R., J. E. Nielsen, and L. Serrano. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 320:369–387.
- [66] Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano. 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* 33:W382–W388.

- [67] Merlo, C., K. A. Dill, and T. R. Weikl. 2005. Φ -values in protein folding kinetics have energetic and structural components. *Proc. Natl. Acad. Sci. USA*. 102:10171–10175.
- [68] Weikl, T. R., and K. A. Dill. 2007. Transition states in protein folding kinetics: The structural interpretation of Φ -values. *J. Mol. Biol.* 365:1578–1586.
- [69] Weikl, T. R., M. Palassini, and K. A. Dill. 2004. Cooperativity in two-state protein folding kinetics. *Protein Sci.* 13:822–829.
- [70] Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics*. 11:1–21.
- [71] Fersht, A. R., and S. Sato. 2004. Φ -value analysis and the nature of protein folding transition states. *Proc. Natl. Acad. Sci. USA*. 101:7976–7981.
- [72] de los Rios, M. A., B. K. Muralidhara, D. Wildes, T. R. Sosnick, S. Marqusee, P. Wittung-Stafshede, K. W. Plaxco, and I. Ruczinski. 2006. On the precision of experimentally determined protein folding rates and Φ -values. *Protein Sci.* 15:553–563.
- [73] Ruczinski, I., T. R. Sosnick, and K. W. Plaxco. 2006. Methods for the accurate estimation of confidence intervals on protein folding Φ -values. *Protein Sci.* 15:2257–2264.
- [74] Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56.
- [75] Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD - Visual Molecular Dynamics. *J. Mol. Graphics*. 14.
- [76] Merritt, E. A., and D. J. Bacon. 1997. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* 277:505–524.

Table 1: Mutational data for the FBP WW domain

mutation	Φ_{exp}	$\Delta G_{N,\text{exp}}$	affected elements
E7A	0.67 ± 0.21	0.52 ± 0.16	hp 1
W8F	0.24 ± 0.03	1.65 ± 0.16	hp 1, hc
T9A	-0.09 ± 0.04	0.93 ± 0.09	hp 1
T9G	0.94 ± 0.20	0.50 ± 0.10	hp 1
Y11A	0.55 ± 0.10	0.63 ± 0.11	hp 1
T13A	-0.03 ± 0.07	0.81 ± 0.17	hp1, hc
T13G	-0.32 ± 0.25	0.58 ± 0.22	hp 1, hc
A14G	0.69 ± 0.28	0.50 ± 0.22	hp 1
D15A	0.82 ± 0.16	0.42 ± 0.09	hp 1
D15G	0.77 ± 0.17	0.39 ± 0.09	hp 1
G16A	1.17 ± 0.22	1.33 ± 0.27	hp 1
T18A	0.93 ± 0.27	0.54 ± 0.17	hp 1
T18G	0.73 ± 0.05	1.14 ± 0.09	hp 1
Y19A	0.11 ± 0.05	0.67 ± 0.13	hp 1, hp 2
Y20F	0.05 ± 0.16	0.68 ± 0.18	hp 1, hp 2, hc
Y21A	0.28 ± 0.02	1.70 ± 0.10	hp 1, hp 2
R24A	0.29 ± 0.09	0.78 ± 0.17	hp 1, hp 2
T25A	0.39 ± 0.04	2.51 ± 0.18	hp 2
T25S	0.27 ± 0.03	1.08 ± 0.09	hp 2
L26A	0.08 ± 0.08	0.56 ± 0.12	hp 2
L26G	0.45 ± 0.04	-1.29 ± 0.10	hp 2
E27A	0.12 ± 0.04	1.02 ± 0.13	hp 2, hc
T29G	0.09 ± 0.02	1.89 ± 0.11	hp 2, hc
W30A	0.19 ± 0.06	0.76 ± 0.14	hp 2, hc
L36A	-0.30 ± 0.16	0.91 ± 0.14	hp 2, hc
L36V	-0.13 ± 0.09	0.53 ± 0.14	hp 2, hc

Experimental Φ -values and stability changes $\Delta G_{N,\text{exp}}$ are from Petrovich et al. [33]. The information on the structural elements affected by the mutations is derived from the contact map shown in Fig. 3. These structural elements are the hairpin 1 (hp 1), hairpin 2 (hp 2), and the small hydrophobic core (hc) of the protein (see text).

Table 2: Experimental and calculated stability changes for mutations of the FBP WW domain that affect several structural elements

mutation	$\Delta G_{N,\text{exp}}$	ΔG_N	ΔG_1	ΔG_2
W8F	1.65 ± 0.16	2.39	0.21	–
T13A	0.81 ± 0.17	0.69	0.22	–
T13G	0.58 ± 0.22	(1.28)	(0.56)	–
Y19A	0.67 ± 0.13	(2.65)	(1.60)	(1.01)
Y20F	0.68 ± 0.18	(−0.76)	(0.31)	(−0.45)
Y21A	1.70 ± 0.10	2.58	0.56	1.42
R24A	0.78 ± 0.17	(−0.23)	(−0.31)	(−0.38)
E27A	1.02 ± 0.13	(0.17)	–	(0.17)
T29G	1.89 ± 0.11	1.47	–	1.14
W30A	0.76 ± 0.14	1.32	–	0.53
L36A	0.91 ± 0.14	0.47	–	−0.30
L36V	0.53 ± 0.14	(0.23)	–	(−0.34)

Experimental data for the stability changes $\Delta G_{N,\text{exp}}$ are from Petrovich et al. [33]. The stability changes ΔG_N , ΔG_1 , and ΔG_2 for the whole protein and hairpin 1 or 2, respectively, have been calculated with the program FOLD-X [65, 66]. For mutations to alanine (A) or glycine (G) and the mutation W8F, native structures for the mutant proteins have been generated by truncation of atoms. For the mutations Y20F and L36V, mutant structures were generated with the program WHAT IF [74]. The wildtype structure used in the calculations is model 1 of the PDB structure 1E0L [35]. To calculate ΔG_1 and ΔG_2 , substructures consisting of the residues 1 to 24 and 15 to 37 of the PDB structure have been used. The FOLD-X calculations have been performed at the ionic strength 150 mM and temperature 283 K of the experiments [33]. Numbers in brackets indicate that the calculated stability changes are not reliable since ΔG_N differs by more than a factor 2 from $\Delta G_{N,\text{exp}}$.

Table 3: Mutational data for the PIN WW domain

mutation	Φ_{exp}	$\Delta G_{N,\text{exp}}$	affected elements
L7A	0.18 ± 0.07	2.06	hc
R14F	0.68 ± 0.11	1.29	hp 1
M15A	0.63 ± 0.14	0.90	hp 1
Y23L	0.64 ± 0.08	1.51	hp 1, hp 2
Y24F	0.52 ± 0.14	0.87	hp 1, hp 2
F25L	0.49 ± 0.08	1.69	hp 1, hp 2
N26D	0.33 ± 0.05	2.13	hp 1, hp 2
T29D	0.30 ± 0.07	1.77	hp 2
A31G	0.44 ± 0.06	1.88	hp 2, hc
W34A	0.36 ± 0.13	1.12	hp 2
K13 κ	0.50 ± 0.05	1.00	hp 1
S16 σ	0.70 ± 0.05	1.39	hp 1
R17 ρ	0.78 ± 0.11	0.74	hp 1
S19 σ	0.83 ± 0.04	2.03	hp 1
H27 η	0.28 ± 0.03	1.77	hp 2
S32 σ	0.51 ± 0.03	1.77	hp 2

Experimental Φ -values and stability changes $\Delta G_{N,\text{exp}}$ for the mutations L7A to W34A are from Jäger et al. [25], and for the amid-to-ester mutants K13 κ to S32 σ from Deechongkit et al. [29]. Here, only mutations with stability change $\Delta G_{N,\text{exp}} > 0.8$ kcal/mol are considered. The structural elements affected by the mutations are assessed from the contact map shown in Fig. 3. These structural elements are the hairpin 1 (hp 1), hairpin 2 (hp 2), and the hydrophobic core (hc) of the protein (see text).

FBP WW domain



PIN WW domain



Figure 1: Native structures of the FBP [35] and the PIN WW domain [36]. The structural representations have been generated with the programs VMD [75] and Raster3D [76].

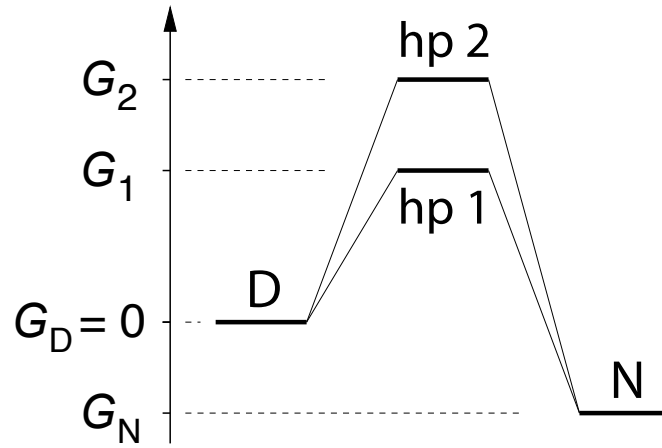


Figure 2: Simple energy landscape of the four-state model for WW domains. The four states are the denatured state D , the native state N , and two transition-state conformations $hp\ 1$ and $hp\ 2$ in which one of the two hairpins is formed. Here, G_N is the free-energy difference between the native state N and the denatured state D , which has the ‘reference free energy’ $G_D = 0$, and G_1 and G_2 are the free energy differences between the transition-state conformations and the denatured state.

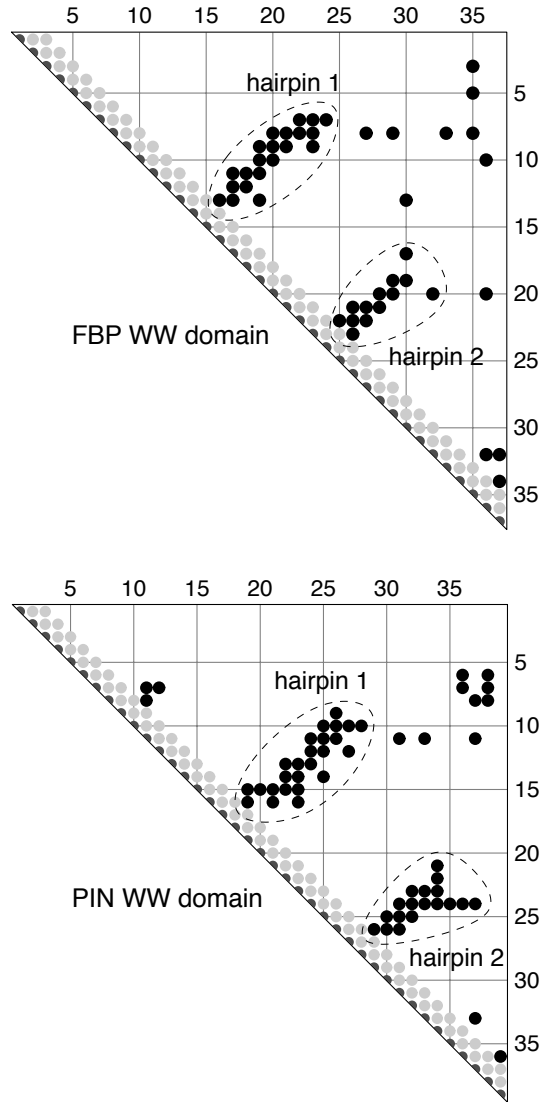


Figure 3: Contact matrices of the FBP and PIN WW domains. A black dot at position (i, j) of a matrix indicates that the residues i and j are in contact. Two residues are defined here to be in contact if the distance between any of their non-hydrogen atoms is smaller than the cutoff distance 4 Å. The hairpins 1 and 2 of the WW domains correspond to clusters of contacts.

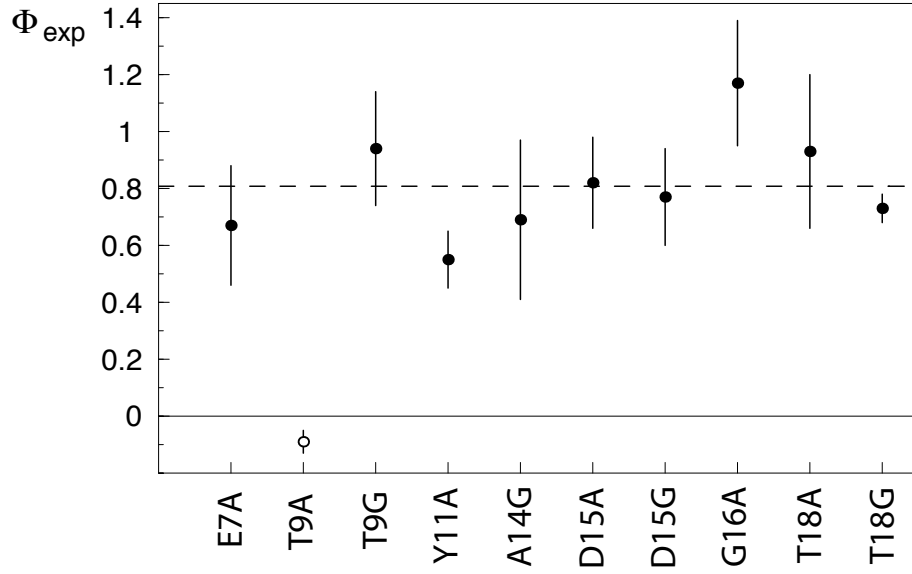


Figure 4: Φ -values for mutations that only affect haipin 1 of the FBP WW domain (see also Table 1). Except for one outlier (open circle for mutation T9A), the Φ -values are centered around the mean value 0.81 ± 0.06 , with deviations mostly within the estimated experimental errors [33].

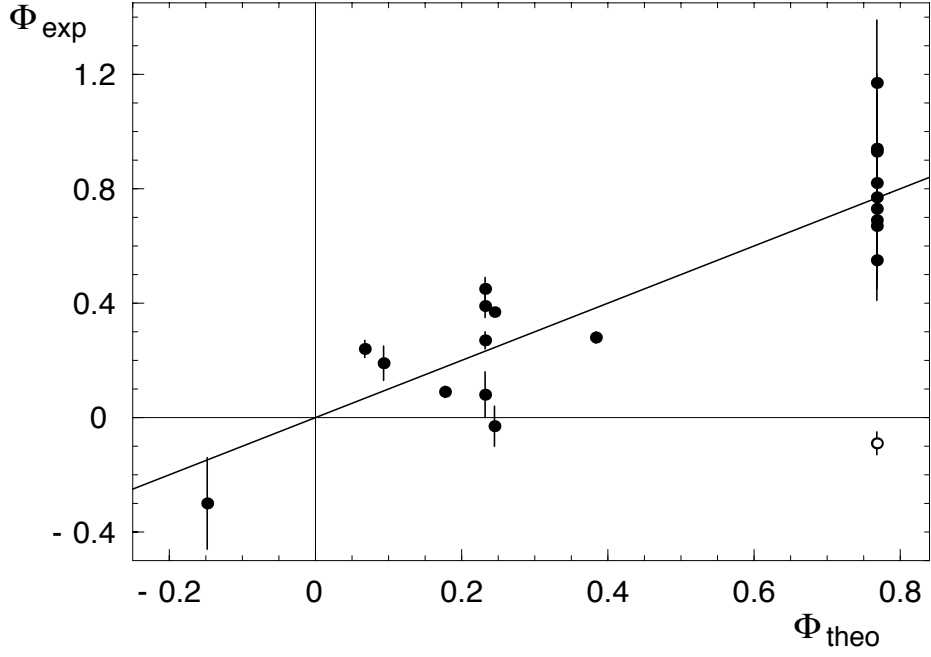


Figure 5: Experimental Φ -values for the FBP WW domain versus theoretical Φ -values obtained from a least-square fit of eq. (3) with the single fit parameter χ_1 . From this fit, we obtain the values $\chi_1 = 0.77 \pm 0.05$ and $\chi_2 = 1 - \chi_1 = 0.23 \pm 0.05$ for the fractions of the two transition-state conformations in which either hairpin 1 or hairpin 2 are formed. The Pearson correlation coefficient between theoretical and experimental Φ -values is $r = 0.90$ if the outlier data point for mutation T9A (open circle) is not considered, and $r = 0.77$ if the outlier is included.

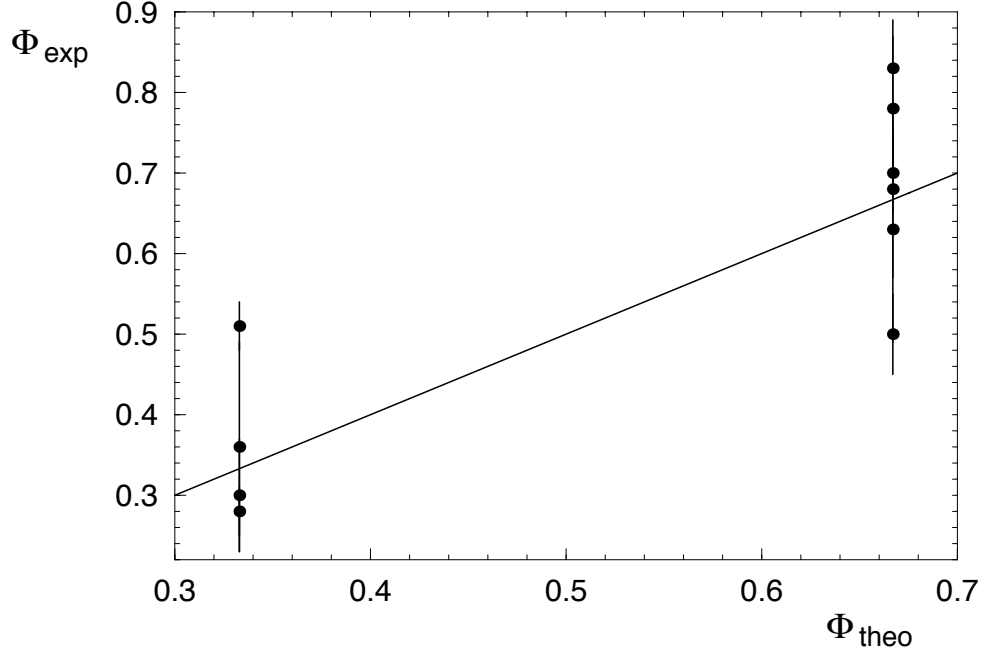


Figure 6: Experimental Φ -values for the PIN WW domain versus theoretical Φ -values obtained from a least-square fit of eq. (3), which results in the values $\chi_1 = 0.67 \pm 0.05$ and $\chi_2 = 1 - \chi_1 = 0.33 \pm 0.05$ for the fractions of the two transition-state conformations. The Pearson correlation coefficient between theoretical and experimental Φ -values is $r = 0.85$.